Signature Recognition System for Student Attendance System in UTP

by

Hafizah bte Abu Bakar

Dissertation submitted in partial fulfilment of
the requirements for the
Bachelor of Technology (Hons)
(Information Technology)

JUNE 2004

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Signature Recognition System for Student Attendance System in UTP

by,

Hafizah bte Abu Bakar

A project dissertation submitted to the
Information Technology Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons.)
(INFORMATION TECHNOLOGY)

Approved by,

(Mr Jale Ahmad)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH PERAK
June 2004

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons

----

HAFIZAH BT ABU BAKAR

ABSTRACT

This paper proposes an off-line signature recognition system for student attendance system in Universiti Teknologi PETRONAS (UTP). In current system, attendance sheet is passed across the class and students are required to signed on the paper. Later, lecturers will check on the paper and mark any empty column. However, lecturers always busy and seldom have time to check each signature. Basically, the system has the ability to imitate humans' capability of recognizing signatures. Thus, it could help lecturers in recognizing students' signatures. The system employs artificial neural networks for recognition and training process. This system is developed mainly using Visual Basic 6.0 and involves four basic steps, which are image acquisition, image pre-processing, and enrolment and verification process. It has two phases, training and recognition. Both process use artificial neural network. The system was satisfactory in all cases where there were two different signatures to be recognized with False Rejection Rate (FRR) for genuine signature is 4% and False Acceptance Rate (FAR) for forged signature is 28%.

ACKNOWLEDGEMENT

BISMILLAH AR-RAHMANI AR-RAHEEM

In the Name of Allah, The Most Compassionate, the Most Merciful

I must thank my GOD who gave me the spirit and strength to work and proceed with this project, going through all the problems and finally able to deliver it.

My sincerest thanks go to my supervisor, Mr. Jale. Thanks for his supports, patience and ideas in assisting me with this project, his never-ending enthusiasm inspires me.

I also would like to express my gratitude to staff of Universiti Teknologi Petronas, especially IT IS committee and lecturers for providing knowledge and resources and willing to help with full commitment. Your supports and commitments really help me.

Special thanks to my housemates, colleagues, and friends for helping me with this project and for always be at my side whenever I need the helps. Thanks for the courage, the passion and the friendships we shared together.

Credits to those who help and guide me on my project via Internet. Thanks for spending their precious time answering all my questions.

Last but not least, deepest gratitude to my parents and family for their enormous material and spiritual supports, without them this work would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

FRR False Acceptance Rate

FAR False Rejection Rate

TAR True Acceptance Rate

TRR True Rejection Rate

UTP Universiti Teknologi PETRONAS

MPL Multi Layer Perception

HMM Hidden Markov Model

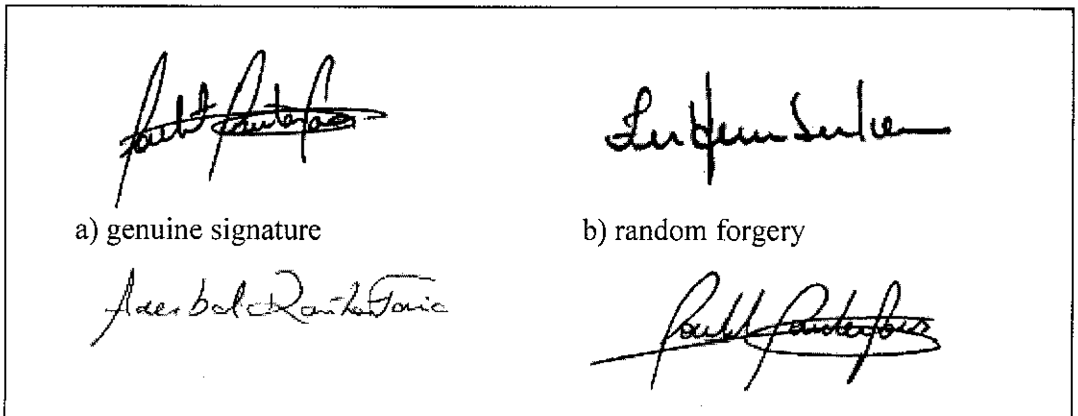
ANN Artificial Neural Network

CHAPTER 1

INTRODUCTION

A signature is any written specimen in a person's own handwriting meant to be used for identification. Signature is a behavioural biometric, it is not based on physiological properties of individual, such as face or iris pattern. Because of this, one's signature may change over time and not unique. However, its wide acceptance and easy implementation make it one of the most famous techniques for identification purposes.

By forgery is meant copying falsifying or altering any kind of written or printed matter for the purpose of defrauding others. There are three different types of signature forgeries, which are random forgeries, simple forgeries and skilled forgeries (Figure 1).



a) genuine signature

b) random forgery

c) simulated simple forgery

d) simulated skilled forgery.

Figure 1.1 Type of forgeries

Random forgeries is produced without knowing either the name of the signer nor the shape of its signature, simple forgeries is produced knowing the name of the signer but without having an example of his signature while skilled forgeries is produced by people who, looking at an original instances of the signature, attempt to imitate it as closely as possible.

1.1 Study on handwriting

Earliest work on handwriting recognition was carried out in the sixties and seventies. The first research works with recognition of isolated hand-printed characters and then followed by the study on the printed character recognition (i.e. OCR). Later, the researchers began to study on the recognition of on-line cursive handwriting. They have achieved successful results in handwriting recognition. Table below shows the summary of recognition studies.

Table 1.1 Summary of recognition studies

	Author	**Recognition rate (%)**
Numeral	Denker et al	86
	Bottou et al.	91.9
	Srihari	89.93
	Lee	99.5
Character	Koutsougeras and Jameel	65-98
	Liou and Yang	88-95
	Srihari	85-93
	Shustorovich	89.40-96.44
Cursive Word	Edelman et al	50
	Lecolinet et al.	53
	Leroux et al.	62
	Chen et al.	64.9-72.3
	Bozinovic et al.	54-72
	Senior and Fallside	78.6
	Simon	86
	Guillevic and Suen	76-98.5
	Bunke et al.	98.27

* (Source: Verma, B. "Recent achievements in off-line handwriting recognition systems")

There are various techniques used for handwriting recognition such as Hidden Markov Models (HMMs) and Neural Network. General approach used in handwriting recognition consists of several steps; image conversion, line separation, skew angles, pre-processing, slant correction and smoothing the chain code contour and lastly word segmentation.

Line separation is accomplished by examining the horizontal histogram profile at the small range of skew angles. Noise is removed in pre-processing stage and later in the word segmentation stage, the word boundaries in the line is located and recognized.

Handwriting recognition has a number of different areas as shown in Figure 1.2. The two major methods of handwriting recognition are online and offline. Both text and signature then can be further divided into recognition, identification and verification

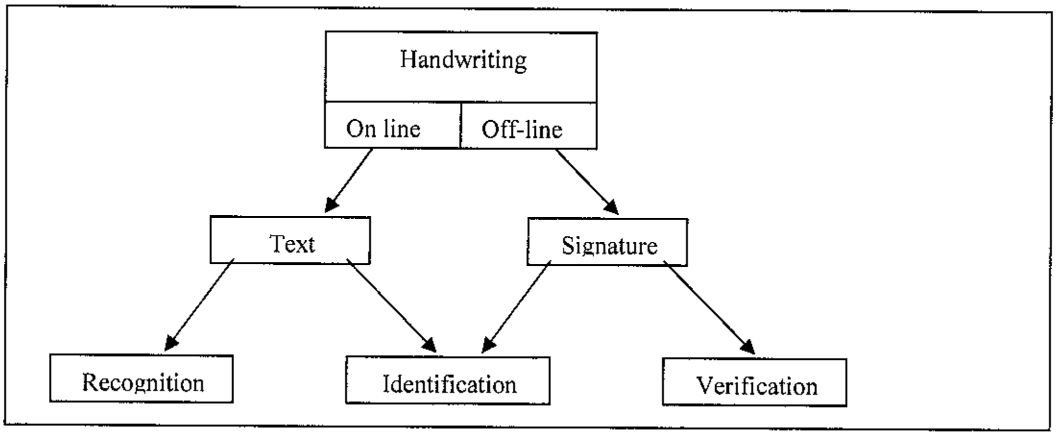


Figure 1.2 Subdivisions of machine handwriting recognition

1.2 Study on signature

The early work on signature verification can be traced back to the 1960s. From that point, many works has been done to improve the techniques.

Signature verification can be categorized into two depending to the way the input is acquired. In off-line recognition system, the signature is acquired as image. No special input device involves. The system mostly relies on image processing and feature extraction. This system widely used in banking transactions and documents. For an on-line system, the input is obtained using pressure-sensitive tablet. Users need to sign on that device for verification. The signature is then, will be analysed according to the shape, speed, stroke, pen pressure and timing information during the act of signing. The system could be used in real time applications such as credit card validation or acquire access to laptop or PDA.

In general, on-line signature verification is more reliable compared to off-line signature verification. This is because the calculation is more accurate and human cannot forge the changes in speed, pressure and timing that occur during the act of signing. However, the usage of off-line system is still widely in used because the cost of implementing it is lower compared to on-line system.

1.3 Problem statement

According to undergraduate handbook, students of UTP are required to attend all lectures, tutorials, laboratories and other academic activities. The minimum attendance requirement is ninety percent (90%) for each course. Failing to do so might cause students to be prohibited from sitting for final examination.

To record the attendance, most lecturers use attendance sheet, which is distributed across the class.

This regulation causes students to present in all classes. However, not every times students could obey the rules. Sometimes when they have other activities to attend and when this situation is inevitable, they will ask their friends to forge the signature.

1.4 Problem identification

Since student is not a professional (skilled) forger, the difference between the forge signature and the genuine one is obvious. However, lecturers always busy and do not have enough time to go through each signature. They would just scan through the list, mark any empty line and then record the attendance for future reference.

Using existing technology, a system is designed to imitate human's capability to read and compare the signatures. Through the system, the signature is digitized and then compared with student's genuine signatures, which are stored in a central database. Through this system, lecturers could detect any random forgeries signatures.

Off-line signature recognition is proposed due to the less number of device involve, consequently reduce the cost of implementation. Students just need to sign on the attendance sheet and later the paper would be used as input.

1.5 Significant of the project

This system could help lecturers in checking the attendance list. It could detect random forgeries, thus making it difficult for students to sign for their friends without looking at the signature first.

System's ability to differentiate between two different signatures could, at least encourage students to attend the lectures.

1.6 Objective and scope of study

In order to develop a system, clearly stated objectives is needed. This is because developer needs to have clear vision of what is being developed. It is common that in every project, the scope is defined to ensure that the project could be completed within the predefined constraints (i.e. time, cost, resources).

1.6.1 Objective

To develop an offline signature recognition system that could imitate human's capability of recognizing signatures.

1.6.2 Scope of study

The system is intended to recognize the signatures and detect the signature that is totally different from the genuine signature.

The limited numbers of sample signatures were collected from students in UTP for training and testing, 14 signatures from each volunteer. To avoid fatigue of hand, five signatures are collected at one time.

Even though in real life application, offline systems have to cope with different pen types, thickness and ink colours, in this research, the process is simplified by using one pen to produce all sample signatures. The signature was collected using black ink pen on white paper. Students are allocated a fixed size box for each signature for easy computation.

Maximum accuracy rate of 80% is expected since the system's functionality is limited to recognize the random forgery.

1.7 Feasibility of the project within the scope and time frame

A system will be designed to recognized and detect the random forgeries. Due to the time constraint (we are given 14 weeks to study and develop the system), system's function will be limited in recognizing the signature. This is to ensure that the system could be completed within the allocated time frame. In other words, it could be used to recognize two different signatures only.

Visual Basic 6.0 is used as main developing software. Visual Basic is chosen because it using easy-to-understand programming language and could provide better interface. Visual Basic has the ability to detect syntax errors as the code typed into the code window. It also provides a number of tools to analyse the programs when they do not work properly such as tools to check the value of variables, step through code and to set breakpoints. Even though Visual Basic is not as powerful as MATLAB in the neural network and statistic area, but its capability is proven. Furthermore, developer's familiarity with Visual Basic environment is an advantage. Scanner and digital camera would be used to acquire the image. Any ordinary scanner with enough resolution could be used for image acquisition. UMAX AstraSlim SE scanner and digital camera model IXUS II from Canon will be used. The scanner and digital camera are always available if needed since these devices are developer's properties.

Another input device is mouse pen. It is a pen-shaped mouse, which would be a great help in acquiring signatures because by using this device there is no image processing involves and this would help reduce a lot of time used to collect signature. However, a few minutes is needed to train volunteers on using this device. It is used to collect signatures for testing purposes only.

CHAPTER 2

LITERATURE REVIEW

2.1 What is neural network?

Neural network is a distributed processor that has natural tendency for storing knowledge based on previous experience and based on knowledge, it would act accordingly. It resembles brain in acquiring knowledge through learning process[16].

2.1.1 Neural network concept

The brain is principally composed of a large number of interconnected specialized cells, called neurons, which can propagate electrochemical signal. The neuron has a branching input structure (the dendrites), a cell body, and a branching output structure (the axon). The axons of one cell connect to the dendrites of another via a synapse. When a neuron is activated, it fires an electrochemical signal along the axon. This signal crosses the synapses to other neurons, which may in turn fire. However, a neuron fires only if the total signal received at the cell body exceeds a certain threshold.

2.1.2 Artificial neural network

Artificial neural networks are made up of many artificial neurons. An artificial neuron is simply an electronically modelled biological neuron. The number of neuron used is depends on the task complexity. It could be as few as three or as many as several thousand. Figure 2.1 below shows the structure of artificial neuron.

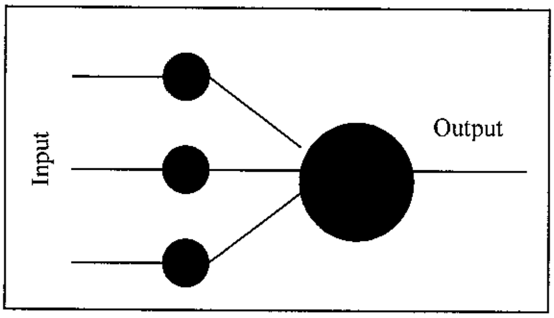


Figure 2.1 The structure of artificial neuron.

Each input into the neuron has its own weight associated with it illustrated by the small circle. A weight is a floating-point number and will be adjusted to train the network. The weights in most neural nets can be both negative and positive, therefore providing inhibitory influences to each input. As each input enters the nucleus (big circle) it is multiplied by its weight. The nucleus then sums all these new input values, which gives the activation. The output signal is determined by the activation and threshold value. This is typically called a step function.

Feed forward network

One way of linking several neurons is by organising the neurons into a design called a feed forward network. It gets its name from the way the neurons in each layer feed their output forward to the next layer it reach the final. Figure 2.2 below shows a very simple feed forward network.

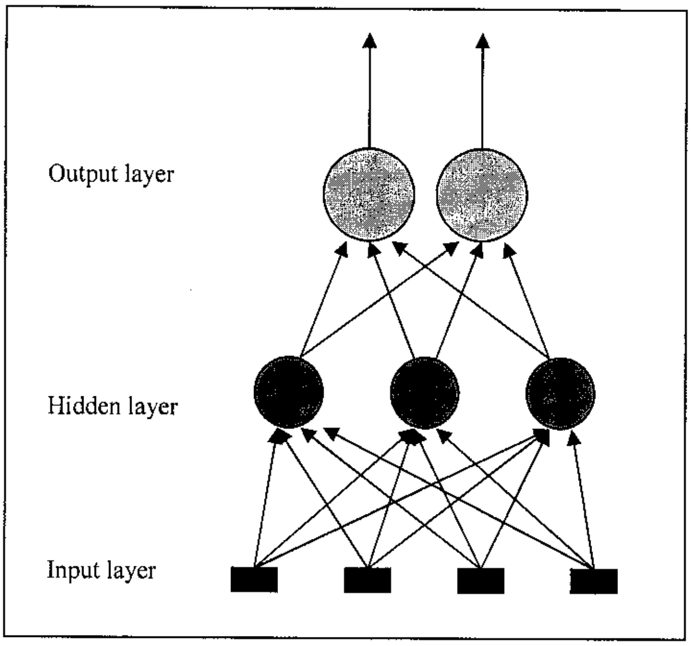


Figure 2.2 Simple feed forward network.

Each input is sent to every neuron in the hidden layer and then each hidden layer's neuron's output is connected to every neuron in the next layer. There can be any number of hidden layers within a feed forward network but one is usually enough.

Generalization

One of the major features of neural networks is their ability to generalize that is to successfully classify patterns that have not been previously presented [10]. The net will learn to key off of significant similarities in the input vectors and ignore irrelevant data. This means that the generalization properties will allow similar inputs as well as noisy inputs to be classified by virtue of their similarity to the training set data. The net generalizes from the taught patterns to include these other similar patterns, which are therefore collectively known as the generalization set. The size of the generalization set is controlled by the diversity of patterns in the training set. The training set should cover the entire expected input space. If the network is inadequately or insufficiently trained on a particular class of input vectors, subsequent identification of members of that class may be unreliable. Fewer hidden nodes force more generalization. This is due to increasing the number of discriminants we have as we increase the number of hidden units [10].

Supervised and unsupervised learning

Once the neural network has been created it needs to be trained. One way of doing this is initialize the neural net with random weights and then feed it a series of inputs. For each configuration output is determined and the weights is adjusted accordingly [17]. This type of training is called supervised learning. In other word, it needs a teacher to tell the network what the desired output should be. The examples of supervised network are simple perceptions, back-propagation, and RBF networks [18].

In an unsupervised net, the network adapts purely in response to its inputs. Such networks can learn to pick out structure in their input [18].

2.2 Application of neural network in signature recognition system

Work done by Xiao and Leedham [15] is based on the Multi-layer perception (MLP). Through the system, the expert forensic first examine the input signature and locates any differences between the input signature and the genuine one by comparing local features before analyse the stability of these features in the genuine samples.

The system consists of three parts: feed forward path, converter and feedback path. The feed-forward path is a modified MLP with two hidden layers and layer 1 is weight shared. The first hidden layer is modified to be weight-shared, to increase the tolerance of writing variance. The converter converts the initial output into the inputs of the feedback path. The feedback path received the input and then produced a feedback signal to the feed-forward path.

By incorporating the feedback path, the false rejection rate is reduced from 15.9% to 9.2%. The strategy used for forgery generation increased the system's ability to distinguish forgeries from 38.9% to 17.0 FAR.

Cemil et al used combination of moment invariant method and artificial neural networks in [4]. For recognition, they used multilayer feed-forward ANN, which consists of 14 input variables, 18 hidden neurons and 30 output variables. To verify the signature, a multilayer feed forward network is also being used. It consists of 14 input variables, 10 hidden neurons and 2 output variables. The system showed 100% success rate by identifying all 30 signatures that it was trained for.

Abbas [10] did an experiment using neural network with four different types of networks. The networks consist of 5600 neurons for input layer, 1 neuron for output layer and variance number of neurons in hidden layer. Network 1 with no neutron in hidden layer, network 2 with 1 neutron in hidden layer, network 3 with 50 neutrons in hidden layer, network 4 with 100 neutrons in hidden layer, using threshold value of 0.5. The result showed that the increasing number of unit in hidden layer does not improve the classification performance and the time needed to train the network was longer. Further, the result proved the fact that fewer hidden layer forces more generalization.

2.3 Feature selection and extraction

Feature selection is the process of choosing input to the pattern recognition system. In [10] Abbas suggested that the features selected should be computationally feasible, lead to a good classification system with few misclassification errors and could reduce the problem data into manageable amount of information without discarding vital information.

Classification rarely performs based on single feature. Often, several features are used to adequately distinguish inputs that belong to different classes. Most common features used are global shape descriptor such as upper envelope, lower envelope, vertical and horizontal projections [1,13], and extended shadow code [7, 12, 14]. These features are used because of the promising result for random forgery detection [12] and simple forgeries [5]. Some researchers used outline of the signature [5], high-pressure regions of signature [5], pixel distribution, signature skeleton [7], and local granulometric size distributions as local shape descriptor [11].

Using extended-shadow-code, the hyper-centre of inertia is evaluated first by considering the silhouette of the signature [14]. Then the whole of signature is translated towards the centre of image area. The projection area (Figure 2.5) is used as shadow mask. The signature is projected onto this projection area (Figure 2.6)[13].

Envelopes

The envelope of a signature is the pair of curves that enclose the signature from above and below. The upper envelope is the outline of the top portion of the signature, and the lower envelope is the outline of the bottom portion. Using this process, most of the pixels in the signature having been ignored as not belonging to one of the envelopes.

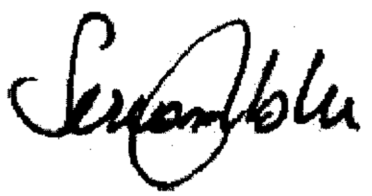


Figure 2.3 Sample of off-line signature

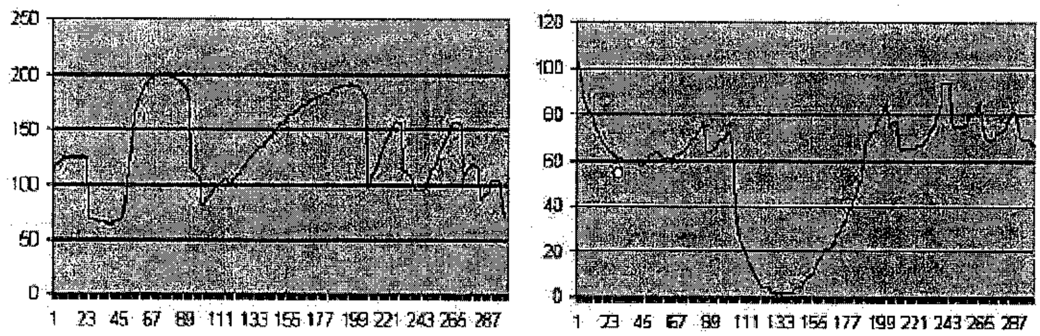


Figure 2.4 Upper and lower envelops of the signature shown in figure 1.3

Extended shadow code

In [14] Robert Sabourin and Ginette Genest described a method of signature verification using a form of shape feature-extraction referred to as an extended-shadow-code, or shadow mask. A shadow mask is a configuration of sampling squares, which cover the entire signature image, each sampling a small area of the signature. A typical shadow mask configuration consists of a constant number of sampling squares, say NxM such boxes, to cover the entire image area. Figure below shows the six projection areas of shadow mask.

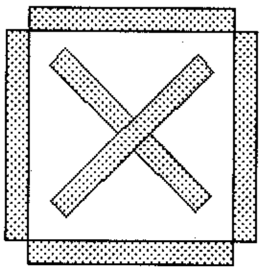


Figure 2.5 Projection areas on shadow mask

The shadow mask consisting of a constant number of sampling squares, where a sampling square consists of top, bottom, right, left, and two diagonal shadow mask bars at 45 and -45 degrees.

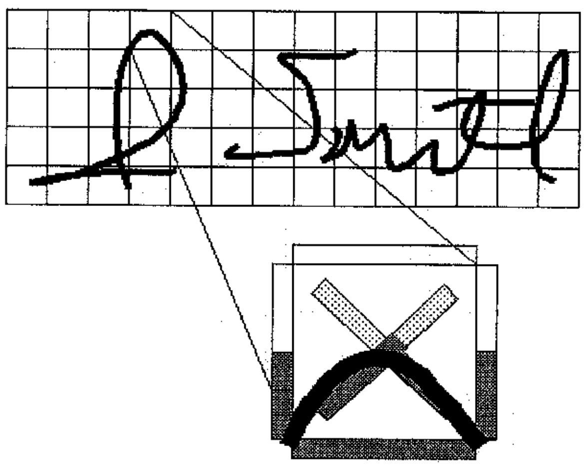


Figure 2.6 Signature with shadow mask

The signature is drawn over the grid of these squares, centred at the hyper centre of inertia. Each shadow calculation projects the signature on each square onto projection areas.

A feature vector is created for each image, each feature being a count of the number of pixels in the shadow that is cast onto each of the bars, cast from the centre outward. Each signature is centred to allow good evaluation of handwriting stability. The hyper-centre of inertia (essentially the centroid, of each signature) is computed to translate the signature to the centre of the image prior to the casting of shadows. By using this way, the signatures of the same person would overlap each other, assuming they were all placed one on top of another, even though they may not line up the same way with the boundaries of the sampling boxes. Each signature was made translation-invariant by translating it according to the hyper-centre of inertia prior to the casting of shadows.

2.4 Problems in offline signature recognition system

Using offline recognition system, skilled forgeries are difficult to detect because lack of suitable features. Unlike offline system, online system is proven more reliable because it uses dynamic features to distinguish between signatures such as signing speed and pen tip pressure [1].

Biometrics.com [19] stated that in online recognition system, the system focused on how the signature was made instead of what the signature looks like. It takes into account the changes in speed, pressure and timing that occur during the act of signing. They believe that there will always be slight variation in signatures but consistency in natural motion creates recognizable patterns.

Due to the difficulties, much of the research effort in the area has been spent on random and simple forgery detection, where deceit is generally obvious. Unlike random forgery detection, skilled forgery detection is a much more difficult task. The fact that no genuine signatures are exactly the same should not be neglected. Thus, a signature recognition and verification system should be able to differentiate between the variation within genuine signatures and the fraud. Offline signature verification involved in skilled forgery detection is still an open research question.

While genuine signatures of the same person may slightly vary, the differences between a forgery and a genuine signature may be imperceptible, which make automatic offline signature verification be a very challenging pattern recognition problem. Besides, the difference in pen widths and unpredictable change in signature's aspect ratio are other difficulties of the problem [1].

It is suggested that in the offline recognition system, the system should classify the signature into three classes; genuine, forgery and uncertain instead of two classes; genuine and forgery. Then, the uncertain signature if left to the human examiner to evaluate the signature [1]. Another suggestion is by using multi-stage system so that system could separate and evaluate the signature stage by stage. The first stage uses outline of signature as a feature to detect the random and simple forgeries while the second stage uses more specific feature, which is high-pressure regions of signature. The system will refuse to decide on the signature with uncertain status, which resulting in rejection. The rejected signature then will proceed to the next stage [5].

2.5 Further development of signature recognition and verification system

In the era of technology, signature recognition is becoming more and more popular. Many works have been done in years by researches to improve the techniques. The early work on signature verification can be traced back to the 1960s. Plamondon and Loretta presented a summary of previous work in 1989, and later in 1994 further summary was made by Leclerc and Plamondon [8].

Mighell et al. was apparently the first to work on offline signature verification system. Feed forward network was used to detect casual forgeries and only one person is used to supply the data for training. The result of the work encouraged McCormack and Brown to extend the previous work. This time six candidates were used. The representation of the data in the Fourier Transform frequency domain as a pre-processing technique for the training and test data is discussed. They found out that the usage of 2D Fourier Transform of normalized signature images resulted in improving the networks ability to generalize, but worsened the false signature acceptance rate. Haar wavelet transforms also was used as method of encoding signature images to achieve data reduction and the result showed that Haar wavelet coefficient data gave the best results compared to a backpropagation network [10].

Most of the remarkable work in the area of random forgery detection was done by Sabourin et al. In [14], Sabourin et al. used extended shadow codes, previously proposed by Burr for handwritten character recognition as a shape feature to detect random forgeries.

Then, the works on offline recognition system expand in conjunction with time. More techniques and algorithm introduced to make better system. The usage of Hidden Markov Model is expanded to recognize the signature pattern [2, 7, 9].

In 1997, Sabourin et al. [11] introduced the usage of local granulometric size distribution to define the local shape descriptor for feature extraction. Later, most of the systems proposed used Artificial Neural Network algorithm [4, 10, 15]. This is because this network is easy to train and learn, and the result is within acceptable range.

Up until now most of the systems proposed, while performing reasonably well on a single forgery, decrease gradually when working with all the categories of forgeries simultaneously [5]. Then, in [5] a serial multi-expert system called Automatic Handwritten Signature Verification System is introduced. It consists of two-class classifier, which would divide task by filtering the signatures. The first stage expert would deal with random and simple forgeries and leave the uncertain signature to second stage expert, which uses specific feature to decide. The result shows that the False Rejection Rate for the system increase 68.76% and False Acceptance Rate for random forgeries is increase 92.31%, simple forgeries, 46.11% and skilled 40.11% compared to the method proposed by Huang and Yan in "Offline signature verification based on metric feature extraction and neural network classification"

CHAPTER 3

METHODOLOGY

3.1 System architecture

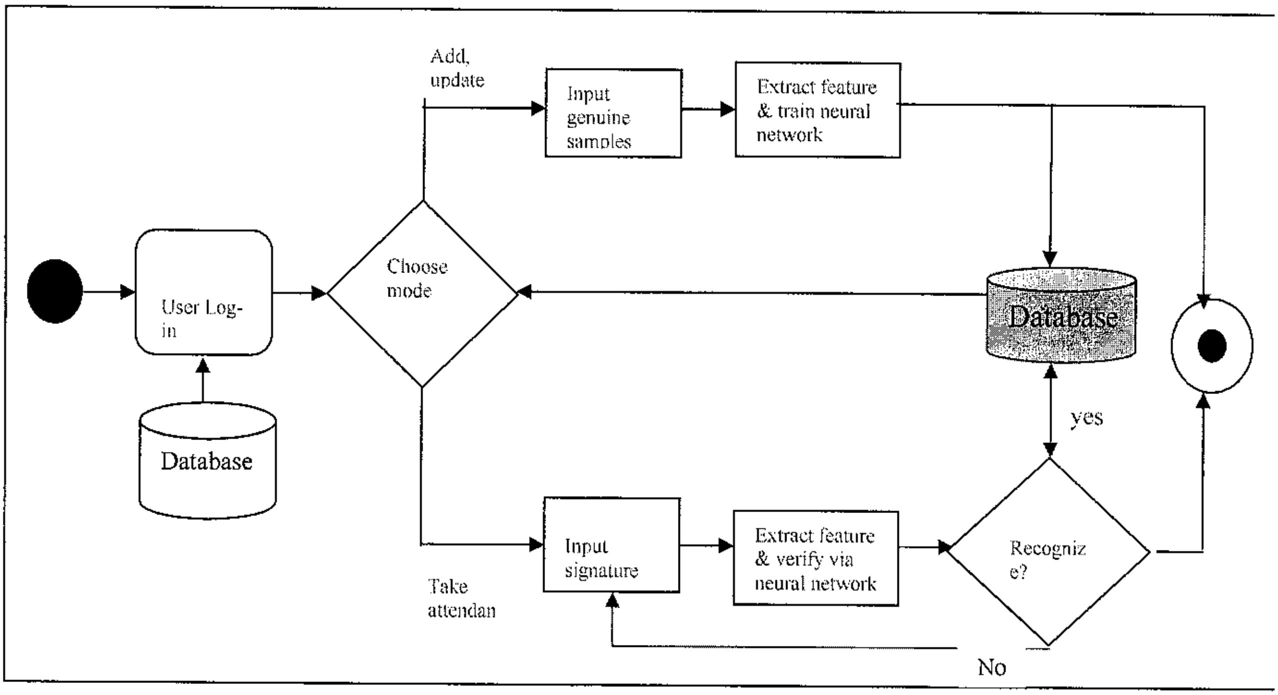


Figure 3.1 System architecture

The system is consists of two modules. The first module is used to accept the input of signature, do comparison with existing signature and if recognized the name of signature owner will be recorded in attendance list as present while the second module is used to add, delete or train the network with new signature. The process in this two modules are same. The input signature is accepted, processed and then the output will be displayed. To use this system, first, the authorized person such as lecturers need to log in to run the application. Users will be prompted with main screen, which required users to users chose the mode, whether the first mode (add, delete or train network), or the second module (take attendance).

If "take attendance" is selected, then user will need to sign on the provided rectangle box and recognition process will take place. If the signature entered is recognized, then user's name will be recorded as present. If the signature is not recognized, then users will be prompted to re-enter the signature.

In second module, the signature entered will be used to learn the pattern of signature. Several numbers of signatures will be needed for the process. In this module few process can be done; add new signature, train or delete existing signatures.

3.2 Project methodology

System life cycle development model is used as the guidelines for system development

3.2.1 Planning phase

In this phase, the purpose was to identify the problem and find possible solutions by performing preliminary investigation or also called as feasibility study. Feasibility study was done to ensure that the project could finish within time frame. The information obtained from Internet, books and journals. From the study, the milestones, project scope and objectives are specified. During this phase, the information collected mostly focused on choosing the project title. After going through some discussion with colleagues and based on the information obtained, developer choose to work on signature recognition using neural network.

3.2.2 Analysis phase

The purpose of this phase was to understand the requirements and build a logical model for the system. The existing application was studied and several aspects were noted. These aspect will be studied further more to determine whether it is feasible to be employed in current project. During this phase, the system architecture is designed and data flow diagram (DFD) is produced to depict the input, process and output of the system.

3.2.3 Design phase

In this phase, all necessary outputs, inputs, interfaces, and processes were identified. All interfaces were designed using Visual Basic. The number of form used is identified, which consists of main page, learning page, user-input page, report page, help page and about page. Based on the designs, the prototype was developed and then showed to supervisor for advice. Further enhancements were made based on the comments.

3.2.4 Implementation phase

During system implementation, the system was constructed. The programs were written, tested, and documented. The objective of this phase was to deliver a completely functioning and documented system. The program was written using Visual Basic. The recognition and learning function was separated using tabbed windows. During this phase, the system evaluation was conducted to determine whether the system operates properly and if costs and benefits were within expectations. Sample signatures were taken for testing. The testing was done using genuine and forged signatures. Based on the result, the FAR and FRR are determined and the accuracy rate is calculated.

3.3 Procedure identification

Several processes are needed to develop the system, which are mage acquisition, pre-processing, feature extraction, enrolment and training and lastly the recognition or validation process.

3.3.1 Image acquisition

Each signature are collected on a 1" x 1.5" rectangle box using the same type of writing tools, which are STABILO point 186, black ink pen and white sheet of paper. It is found that insignificant number of noise would be obtained if the process were using black pen and plain white paper. However, the noise still exist due to the quality of paper and ink used and also poor illumination (if digital camera is used).

To simplify the process of scanning, a group of 70 signatures are collected on a single sheet paper. Limited number of sample signatures was collected from 10 volunteers (students of UTP) for training and testing. 14 signatures were collected for each subject. To avoid fatigue of the hand and the possibility of distortion, only five signatures will be collected at a time.

Signature image could be acquired using any digital devices such as digital camera and scanner. If digital camera is used, the process should take place in a room with adequate lighting to ensure that no blaze of light and consequently ensuring that the intended image is captured. The blaze could cause the paper appears blank when the image is captured.

For testing purposes, the signatures were taken using mouse pen. Using input from mouse would reduce time spent on image processing. However, the usage of mouse input for training is not efficient because quite large number of signatures is collected per volunteer and the process of repetitively signing on tablet is not an easy task.

3.3.2 Pre-processing

Current technology devices mostly used colour scanning while techniques presented in this study are based on grey scale images, therefore, scanned or captured colour images are initially converted to greyscale.

Image scanning device may cause noise to a signature image. Another source of noise may be speckled paper background on which the signature is signed. This noise could thwart the feature extraction process thus, could affect the end result. This noise needs to be removed. To remove the noise and remove it from the background, Adobe Photoshop, the image processing software was used. Through this software, the process of image processing would be much easier.

At first, the sheet of papers containing sample of signatures was scanned using AstraSlim SE scanner. The contrast and brightness was adjusted to make the signatures stand out while the rest of the settings are set to default. Then, using Adobe Photoshop, the RGB image was converted into monochrome image to make it contains only black and white pixels.

The system is designed to accept input of signatures one by one, while the sample of signatures were collected in a group. To conform to the requirement, each signature was cut and saved in different files. Later the images will be used as input for training.

Using off-line signature, the size of signature might vary according to the size of available surface, even the same person may use different size of signature. There is no reliable way of controlling the size of signature. To eliminate the size differences and obtain a standard signature size, the normalization is used. In this study, normalized size for all signatures is used. During normalization process, the aspect ratio of width and height of signature is kept intact. The process is done using Adobe Photoshop.

3.3.3 Feature extraction

The pixel distribution feature called extended-shadow-code is used for feature extraction. This is because the study done shows promising results for random forgery identification. Through this technique, the top, bottom, left and right borders of signature are recognized. Then each black pixel in the input image is then recognized and mapped on the grid for further recognition process.

3.3.4 Enrolment / Training

During enrolment to the system, a user supplied a number of reference signatures, which then are used for calculations describing the variation within reference signatures. For training, the signatures were collected in a fixed-size rectangle box, using the same writing devices, which were STABILO point 186, black ink pen and plain white papers.

3.3.5 Comparison

The signature will be compared to the reference signature for recognition and then displayed the result whether the signature is similar or not. A two-layer neural network with no hidden layer is used to learn and recognize the patterns. The hand-drawn image is digitized onto a grid of input neurons. Each possible answer is represented by a single output neuron. Every input neuron is linked directly to every output neuron. As in most neural networks, the data (or programming) is encoded in the links between neurons. If a link between an input neuron and an output neuron is positive, that means that if the input is on then the total score for that output neuron is increased by a small amount. If the link is negative, then it follows that if that input is on, the corresponding output has its score reduced by an amount. The output neuron with the highest score is considered as the best match pair.

3.4 Tools

To develop this project, some tools are required including the usage of software and hardware. For hardware, a workstation that is capable to run and execute the program is required. Minimum requirement for workstation would be Intel Pentium MMX 166Mhz processor, but for better performance, higher version of processor would be good.

Scanner and digital camera would be used to acquire the image. Any ordinary scanner with enough resolution could be used for image acquisition. For this process, UMAX AstraSlim SE scanner and digital camera model IXUS II from Canon were used. Image taken using digital camera showed poor quality cause by poor illumination and light flare problem while scanning process generate noises to images. Scanner is a better choice since it could eliminate the effects of poor illumination and the noise problem could be solve using image processing software.

For all signatures collected, the same writing devices would be used. To simplify the process of pre-processing later, the signatures were collected using standardized devices, which are STABILO point 186, black ink pen and plain white paper. Another way of collecting signatures was via pen-shaped mouse called mouse pen. This type of mouse is usually used by online signature recognition system because it could take signatures online and make use the advantage of taking online signature as a feature. However, for demonstration and testing purposes, we would use this mouse, which is GENIUS e-pen. The usage of e-pen could reduce time taken to do image processing.

Visual Basic 6.0 is used as main developing software. Visual Basic is chosen because it using easy-to-understand programming language and could provide better interface. Visual Basic has the ability to detect syntax errors as the code typed into the code window. It also provides a number of tools to analyse the programs when they do not work properly such as tools to check the value of variables, step through code and to set breakpoints. Even though Visual Basic is not as powerful as MATLAB in the neural network and statistic area, but its capability is proven. Furthermore, developer's familiarity with Visual Basic environment is an advantage.

CHAPTER 4

RESULTS AND DISCUSSION

In order to test the accuracy of the system, a system testing has been conducted after the main function of the system fully worked. The test data was obtained using mouse pen and involved five volunteers. This section is to discuss the results obtained and reasons behind it.

4.1 Testing procedure

The program allows users to choose the input method; either load the signature image or sign on the provided box, using a mouse pen. Mouse pen is a pen-shaped mouse, which is used for easy writing.

For signature image, it needs to be processed first. The phase of image pre-processing was done using image processing software, Adobe Photoshop. The pre-processing phase includes normalize the image size, remove noise and change it into readable format.

Signatures used for testing the recognition system were obtained using the same way as for training. The recognition system is tested using 10 signatures; five imitations and five true signatures. The test signature is loaded into the image box, digitized and mapped into grids before being recognized. Once the recognized button is clicked, the system would calculate the weight, based on the training data, the percentage of each signature in database is calculated and the signature with highest percentage is considered as the recognized signature.

For signatures obtained using mouse, volunteers were trained to make them comfortable on using it. This is because unfamiliarity could affect their signature. They were given about 15 minutes to make themselves comfortable.

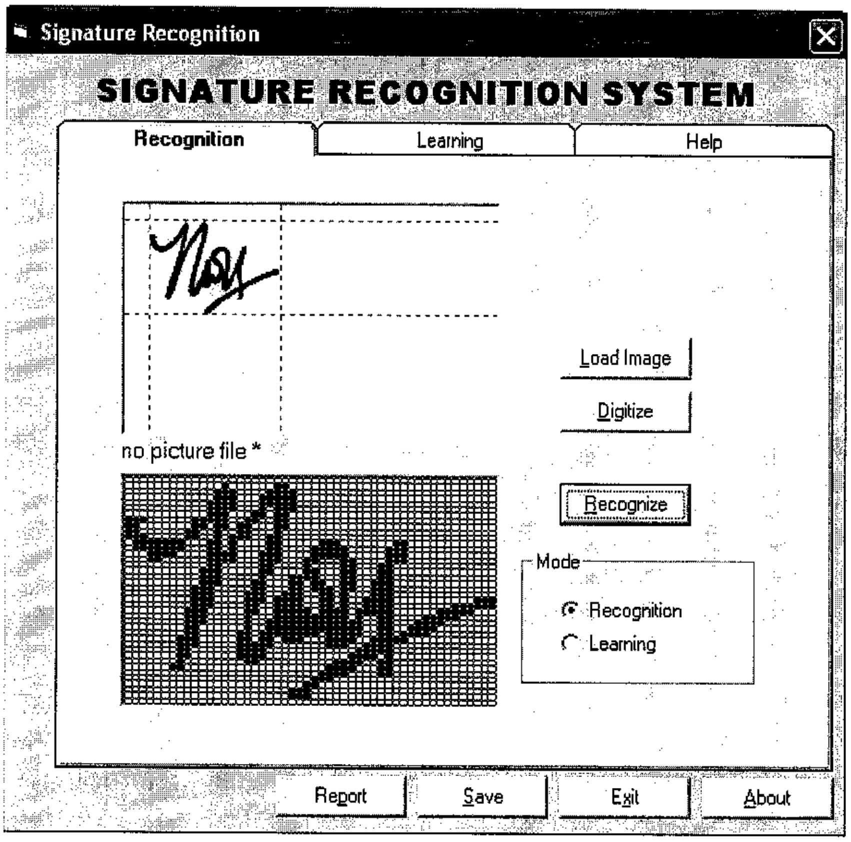


Figure 4.1 Interface for training and testing

4.2 Results of testing

The objective of the testing is to find the extent of recognition accuracy. The testing was done using the testing procedure mentioned above. The first testing involved the genuine signatures stored in database and the signatures taken from the owner of the signature. The result is shown in the table below.

Table 4.1 Testing result using genuine signatures against signatures in database

Sample	Sample 1	Sample2	Sample 3	Sample 4	Sample 5
Normah	X	X	X	X	X
Hafizah	X	X	X	X	X
Halwani	X	X	0	X	X
Izwa	X	X	X	X	X
Faradila	X	X	X	X	X

X = Recognized 0 = Not recognized

The result showed that most of the signatures taken the true owner is recognized successfully by the system. This is because the signature's shape was taken as the feature and every time the data is trained, the shape will evolve. So, when the new input is entered, if it comes from the owner of the genuine signature, then the signature is more likely to be recognized even though the signature differs slightly.

The usage of the shape as the feature makes the signatures is easily recognized. However, the drawback of this feature is that even the signature with the slightly same shape would also be recognized. This is why this program is not suitable for signature verification. If the forgery to be detected is random, which means the forger does not know the name of the signature owner or never see the genuine signature, then the program could detect it.

The second test involved the genuine signature and the forged signature. However, since this application is intended to recognized and detect the random and simple forgeries (student is not a skilled forger), then the forged signature is taken with the assumption that the forgers never see the genuine signature. The summary of the result is shown in table below.

Table 4.2 testing result using forged signatures against signatures in database

Sample	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Normah	X	X	0	X	X
Hafizah	0	X	0	X	0
Halwani	0	0	0	0	0
Faradila	0	0	0	0	0
Izwa	0	0	0	0	X

X = Recognized 0 = Not recognized

The result showed that even the randomly forged signatures could be recognized as authorized. One of the reasons is that the signature owner used her own name as the signature. As example Normah signed as  . This kind of signature can be easily forged even if the forger does not know the actual shape of signature, but by guessing the forger would success.

In evaluating the system, there are two important factors to be considered, which are the FRR of genuine signatures and the false acceptance rate FAR of forgery signatures. The system showed 4% of FRR and 28% for FAR.

The FRR is low because the feature used for comparison accept large variation in signature, thus signature with slight different also would be recognized. FRR and FAR is inversed, so low in FRR would cause high in FAR. However, in the case of signature recognition, 28% of FAR is not normal. Considering the number of sample taken for testing, this is acceptable because the number of samples taken could affect the rate calculated. With only five samples taken for each signature, small number of FAR could give big impact to the FAR rate. As explained earlier, the main objective of this system is to recognize the signature. Hence, the objective is reached with 4% of FRR.

The accuracy rate is the correct recognition obtained divided by the total number of testing, depicts by the formula below.

$$\textbf{Accuracy} = \frac{\textbf{TAR} + \textbf{TRR}}{\textbf{TAR} + \textbf{TRR} + \textbf{FAR} + \textbf{FRR}}$$

Thus, the accuracy rate for the system is 84%.

4.3 Problem faced in the project

The most difficult task is to process the image without using image –processing software. Some algorithms could be used for the task, however Visual Basic cannot handle the pre-processing task as well as MATLAB could. So, the image cannot be loaded into the system without being edited or processed first. In order to do so, Adobe Photoshop is used to overcome the drawback. It takes time because the task need o be done manually and one-by-one. Another way to solve the problem is by allowing direct input via mouse. This task could be done easily using Visual Basic.

When Visual Basic is chosen as main developing software, its lack of statistic ability is acknowledged. The graph showing the percentage or other statistics regarding the signature's difference is difficult to produce using Visual Basic. It is possible but would involve complex coding and time constraint not permits it. So, table showing the percentage rate of the recognized signature is used instead of graph.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

The scope of the project is to recognize the signatures. In order to do so the process of implementation has been implemented using simple signature recognition system, which could recognize the signature and leave the decision on whether the signature is forge on human decision. The process involves training and recognizing the signatures using simple neural network.

5.1 System's limitation

Since system is using grid to represent pixel in signatures, it cannot handle complex signatures or signature with many loops. The dimension of grid box limits system ability. When complex signature is mapped onto the grid, the pixel represented might distort the original shape of signature because not each pixel could be mapped onto the grid. Figure below depict the situation. The curve of the signature is not presented exactly as the original signature and the dimension of the grid make the digitize signature differs from original signature.

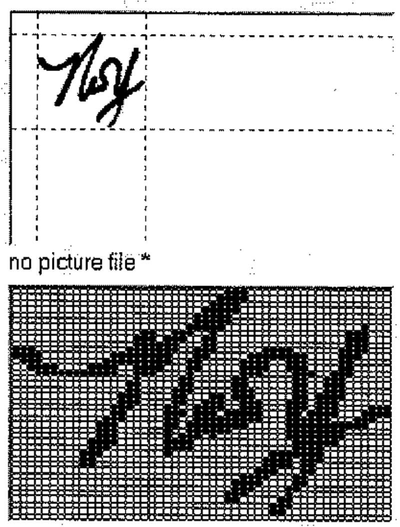


Figure 5.1 Actual image and the image represented in grid

The train data is stored in database that keeps the information of the pixel mapped onto the grid. System then compares and calculates the information with the signature to be recognized. However, the grid concept again limits its ability. The data stored will intersect each time the network is trained. Over trained could cause the system unable to detect variant in signature but signatures that has great variation could cause the signature lost its shape. Figure below shows the information stored for figure 5.1 and notice the differences its shape.

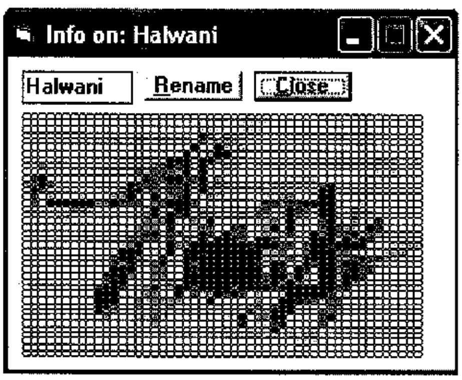


Figure 5.2 The information stored for signature in figure 5.1

The signatures, which have slight similarities in pixel, could be mistaken by the system as the same signature. This is why the system is intended to recognized different signatures, not to detect the forgeries.

The information for each signature is stored using notepad as database. This database is efficient for limited number of signatures, in other word the system is applicable for small applications which not using large database.

5.2 Future recommendation and enhancements

The result showed some interesting results regarding the ability of neural networks to recognize signatures. However, the accuracy rate is still low, and larger training and test data are needed to be able to get reliable statistics rates of the system. The usage of neural network without hidden layer works well if task clearly indicate that the class is linearly separable, which means to distinguish two different signatures. For future enhancement where system could identify and detect the forgeries, hidden layer should be added accordingly. Nevertheless, the fact that time needed to train the network with the larger number of hidden units is longer than the network with zero or one hidden unit should be considered. This is due to the extra calculations, which were required for all the extra connection and unit added

As mentioned before, for more reliable result, a large number of signatures are needed for training purposes. However, it is impracticable to ask client to produce such number. A method that could generate variation of signature within acceptable variance range should be produce. The method could use some techniques such as skewing, rotating and stretching the signature. The usage of notepad as database should be replace by suitable database, such as oracle or MS SQL Server. As the number of students expands, the training data would increase and keeping large information in notepad is inefficient and not secure. Furthermore, the usage of notepad could limit the application of recognition system in small system. By expanding it using Oracle or SQL Server, the system could be employed in complex or crucial system since these types of database emphasize on the security aspect.

Unlike biometric recognition, a person's signature changes over their lifetime and may vary from time to time. For training data, the signatures are taken in three different days to avoid distortion and to ensure that the signatures vary naturally. Signature, just like handwriting changes when human's mood changes. This means the signature should be taken periodically and continuously updated to ensure that the signatures taken reflect the variation in genuine signatures. In real application, the system should be able to adapt to signature evolution. The usage of grid should be avoided because it only limits the system's ability to expand.

In this project, the samples of genuine signatures were collected in a rectangle-bounding box using standardized writing tools. However, in real world student would use different ink colour pen and the signatures sometimes are signed beyond the allocated box. Hence, extra function, which could filter the colour, resize and remove the background image is needed.

This system is developed using Visual Basic, which is not very good in neural network, recognition and comparison between data compared to MATLAB. To enhance and expand its ability in recognizing and validating signature, the system could be upgraded using MATLAB, which its ability is proven in this area.

5.3 Conclusion

Works in signature recognition technology is still ongoing. Online recognition system would always give higher accuracy rate compared to offline recognition system. However, researchers still have interest in offline recognition system and try in every possible way to introduce new way and techniques that could increase its performance. This is because the cost of implementing it is lower with fewer devices used. If the offline recognition system's ability could be increased and more reliable, the digital security industry would cherish with more security application is introduced in lower cost. Hence, the aspiration of implementing the signature recognition and verification system in student attendance system would become true.

Generally, there are several causes that make the system fail to recognize the signature was due to poor image quality, and high similarity between two signatures. Using additional features in input data set and adding the relevant number of hidden layer could increase the recognition ability

REFERENCES

[1] A.A. Kholmatov, 2003, *Biometric Identity Verification Using On-Line & Off-line Signature Verification*, MSc. Thesis, Sabanci University

[2] A. Hershowitz, A. Schwartzman, 1996, *Signature Verification Using Hidden Markov Models: Observations and Alternatives*

[3] B. Verma, M. Blumenstein & S. Kulkarni, *Recent achievements in off-line handwriting recognition systems*

[4] Cemil OZ, Fikret Ercal, Zafer Demir. 2003, "Signature Recognition and Verification with ANN", http://www.eleco.emo.org.tr/ELECO2003/csession/C5-02.pdf

[5] C. Sansone, M. Vento, 2000, "Signature Verification: Increasing Performance by a Multi-Stage System," *Pattern Analysis & Applications vol. 3*, p.169-181

[6] *D. Kalenova, 2003,* "Personal Authentication Using Signature Recognition" *www.it.lut.fi/kurssit/03-04/010970000/seminars/Kalenova.pdf*

[7] E.J.R. Justino, F. Bortolozzi, R. Sabourin, 2001 "Off-line signature verification using HMM for random, simple and skilled forgeries," *Proceedings of the Sixth International Conference on Document Analysis and Recognition*, p. 1031–1034.

[8] K. Zhang, I. Pratikakis, J. Cornelia, E. Nyssen, 2000, "Using Landmarks to Establish a Point-to-point Correspondence between Signatures," *Pattern Analysis & Applications vol. 3*, p.69-75

[9] M.E. Munich, 1995, *Application of Hidden Markov Model to Signature Verification*

[10] R. Abbas, 1994, *A Prototype System for Off-line Signature Verification Using Multilayered Feedforward Neural Networks,* MSc. Thesis, RMIT

[11] R. Sabourin, G. Genest, F. J. Prêteux, 1997, "Off-Line Signature Verification by Local Granulometric Size Distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **19** **(9)**: 976-988

[12] J. R. Parker, L. Bateman, M. Baumback. "Scale Effects in Shadow Masks for Signature Verification," *Proceedings of the International Conference on Artificial Intelligence, 2002 p. 91-95*

[13] J.R. Parker, Brad Behm, "Signature Verification Using Stroke Histograms," *Proceedings of the International Conference on Artificial Intelligence, 2002,p87*-90

[14] R. Sabourin, M. Cheriet, and G. Genest, "An Extended Shadow Code Based Approach for Off-Line Signature Verification," *Proceedings of ICDAR□93, Tsukuba, Japan, 1993. Pp. 1-5.*

[15] Xuhong Xiao and Graham Leedham, *Signature Verification by Neural Networks with Selective Attention and A Small Training Set*

[16] B.D. Ripley. 1996, *Pattern recognition and neural network,* UK, Cambridge University Press

[17] http://www.statsoftinc.com/textbook/stneunet.html/

[18] www.ai-junkie.com

[19] www.cs.stir.ac.uk

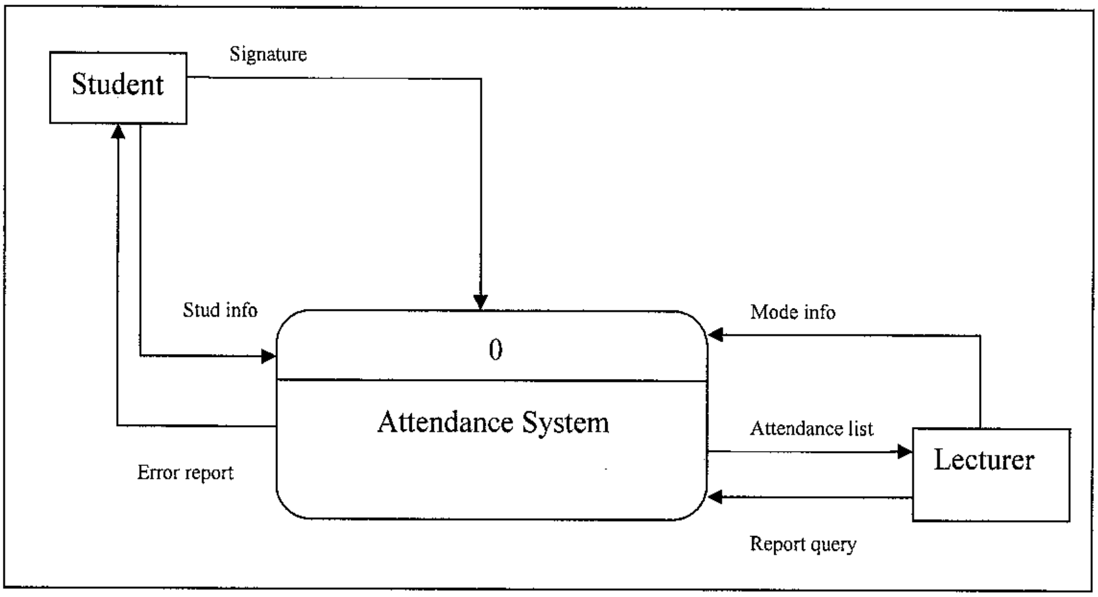
[20] www.biometrics.com

[21] http://vv.carleton.ca/~neil

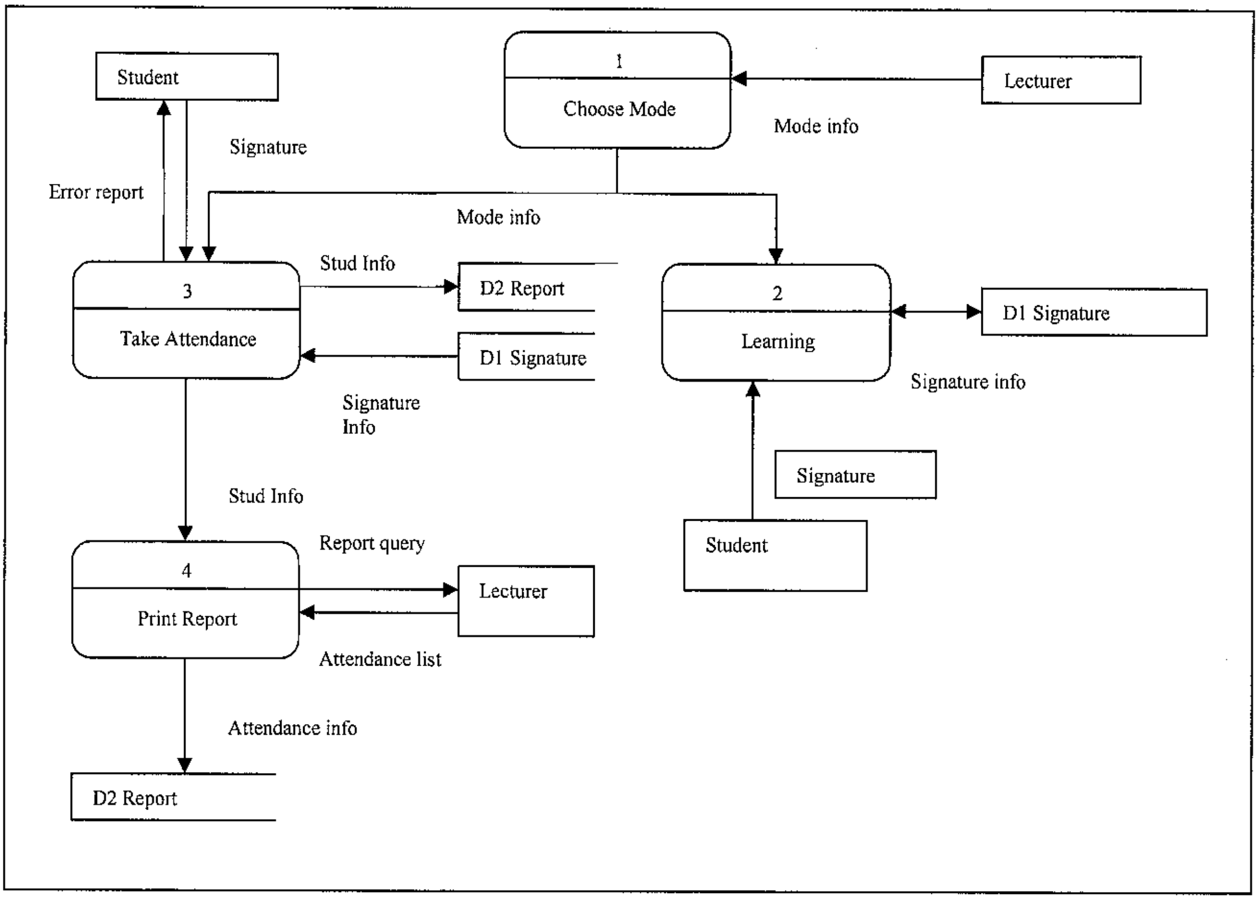
[22] http://home.eunet.no/~khunn/papers/2039.html

APPENDICES

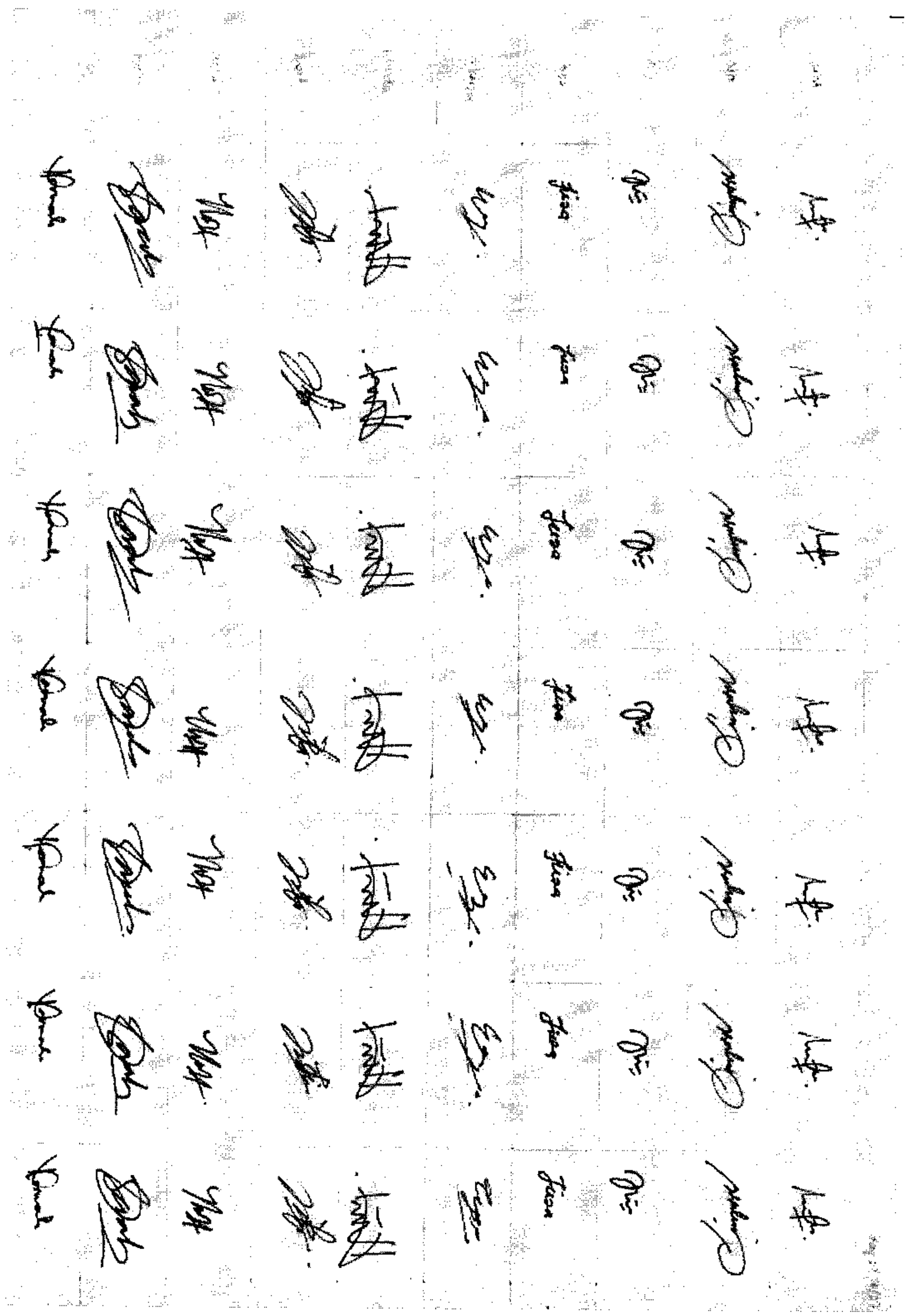
Appendix I Context Diagram for the Student Attendance System



Appendix II Diagram 1 of student attendance system



Appendix III Sample signatures 1



Appendix IV Sample signatures 2

Appendix V Sample signatures 3

Appendix VI Test score sheet for forged and genuine signatures

Sample	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Normah					
Hafizah					
Halwani					
Faradila					
Izwa					